4.4.2 String Similarity

- *4.4.2.1 Explanation of the Subject*

Recommendation 2 states that:

> *Strings must not be confusingly similar to an existing top-level domain or a Reserved Name.*

As implemented in the AGB, in Module 2 it describes string similarity reviews that test "Whether the applied-for gTLD string is so similar to other strings that it would create a probability of user confusion."

This review involves a preliminary comparison of each applied-for gTLD string against existing TLDs, Reserved Names (see subsection 2.2.1.2), and other applied-for strings. The objective of this review is to prevent user confusion and loss of confidence in the DNS resulting from delegation of similar strings. (In the AGB "similar" means strings so similar that they create a probability of user confusion if more than one of the strings is delegated into the root zone.) The visual similarity check that occurs during Initial Evaluation is intended to augment the String Confusion objection (Module 3, Dispute Resolution Procedures) that addresses all types of similarity. A String Similarity Panel conducts this review.

The Panel uses the following standard when determining string confusion:

> *String confusion exists where a string so nearly resembles another visually that it is likely to deceive or cause confusion. For the likelihood of confusion to exist, it must be probable, not merely possible that confusion will arise in the mind of the average, reasonable Internet user. Mere association, in the sense that the string brings another string to mind, is insufficient to find a likelihood of confusion.[1]*

The panel assesses similarities that would lead to user confusion in four sets of circumstances when comparing:

- Applied-for gTLD strings against existing TLDs and reserved names;
- Applied-for gTLD strings against other applied-for gTLD strings;
- Applied-for gTLD strings against strings requested as IDN ccTLDs; and
- Applied-for 2-character IDN gTLD strings against:
  - Every other single character.
  - Any other 2-character ASCII string (to protect possible future ccTLD delegations).

---

[1] Ibid

In addition to the above reviews, an applied-for gTLD string that is a 2-character IDN string is reviewed by the String Similarity Panel for visual similarity to:

a) Any one-character label (in any script), and
b) Any possible two-character ASCII combination.
An applied-for gTLD string that is found to be too similar to
a) or b) above will not pass this review.

The AGB notes:

> *The String Similarity Panel is informed in part by an algorithmic score for the visual similarity between each applied-for string and each of other existing and applied for TLDs and reserved names. The score will provide one objective measure for consideration by the panel, as part of the process of identifying strings likely to result in user confusion. In general, applicants should expect that a higher visual similarity score suggests a higher probability that the application will not pass the String Similarity review.*

> *The panel will also take into account variant characters, as defined in any relevant language table, in its determinations. For example, strings that are not visually similar but are determined to be variant TLD strings based on an IDN table would be placed in a contention set. Variant TLD strings that are listed as part of the application will also be subject to the string similarity analysis.*

> *The panel will examine all the algorithm data and perform its own review of similarities between strings and whether they rise to the level of string confusion. In cases of strings in scripts not yet supported by the algorithm, the panel's assessment process is entirely manual.*

An application that fails the String Similarity review due to similarity to an existing TLD will not pass the Initial Evaluation, and no further reviews will be available. Where an application does not pass the String Similarity review, the applicant will be notified as soon as the review is completed.

An application for a string that is found too similar to another applied-for gTLD string will be placed in a contention set. An application that passes the String Similarity review is still subject to objection by an existing TLD operator or by another gTLD applicant in the current application round.

In addition, applied-for gTLD strings are reviewed during the String Similarity review to determine whether they are similar to a Reserved Name. An application for a gTLD string that is identified as too similar to a Reserved Name will not pass this review.

String Similarity results for the new gTLD applications were published on 26 February 2013.[2]

For those cases of contention that are not resolved through CPE or voluntary agreement, auction is the tie-breaker method of last resort.   An auction of two or more applications within a contention set proceeds as an
ascending-clock auction as described in section 4.3.1 of the AGB.

- *4.4.2.2 Questions and Concerns Related to Subject*

  The DG noted several issues relating to string similarity and auctions. In particular, the DG wondered whether string contention mechanisms were effective in resolving contention. They noted that in order to determine effectiveness, a definition of success may be required.

  In addition, the DG asked whether string similarity resolution methods could be improved or substituted for new mechanisms, such as allowing for string changes or for the substitution of alternate strings. According to the current methodology for reviewing string similarity there is no option for the applicant to alter the applied for string in response to concerns about similarity with existing or other applied for strings. Thus, if a string is rejected due to issues of similarity, an applicant would have to submit a new application for an alternate string, which would have to occur in a subsequent round. The DG also asked whether string contention results were consistent and effective in preventing consumer confusion.

  Moreover, the results of the string similarity review were released two weeks before the deadline to file a String Confusion Objection, so parties who wished to file a String Confusion Objection based on the results of the String Similarity Review (i.e., create contention where the String Similarity Review did not) had a very limited amount of time to prepare an objection. The delayed String Similarity results in this round were caused by the high volume of unique strings, but for future rounds, consideration should be given to how to best position the relative timing of these two processes, taking into consideration unknown factors such as the volume of unique strings.

  Regarding the results, many members in the community, including the DG, the GAC, and the ALAC, raised concerns regarding the similarity of singulars and plurals. As the guidance provided on what constituted confusing similarity in this application round did not provide this level of detail, the standards for confusion may benefit from further refinement for future application rounds.

  With respect to auctions the DG questioned whether additional analysis should be conducted to determine whether auctions are the right mechanism of last resort. They noted that this may require defining the ideal characteristics of a mechanism of last resort.

---

[2] ICANN. (26 February 2013) New gTLD Program: String Similarity Contention Sets. Retrieved from
http://newgtlds.icann.org/en/announcements-and-media/announcement-26feb13-en

- *4.4.2.3 Relevant Guidance*

    o   Recommendation 2

- *4.4.2.4 Rationale for Policy Development*

With respect to the questions and potential issues raised by the DG, it may be useful for ICANN to collect data concerning the results of the string similarity reviews that were conducted. This could be in the form of a survey to the ICANN community. As the results were perceived to be inconsistent by the DG and others, a potential PDP-WG may want to consider providing clearer definitions around what constitutes string similarity to hopefully reduce the possibility of reaching inconsistent evaluation, or even the perception of inconsistency. Specifically, the topics of plurals and the exploration of different ways to resolve string contention have been identified as likely requiring policy development..

With respect to auctions, a potential PDP-WG on New gTLD Subsequent Procedures could consider whether to define "mechanism of last resort" to help determine whether auctions fit the definition and/or whether there are other mechanisms that could be considered.